

# Penyetaraan (*Equating*) Ujian Akhir Sekolah Berstandar Nasional (UASBN) Dengan Teori Tes Klasik

Arniati Prasedyawati Herkusumo  
Peneliti pada Pusat Penilaian Pendidikan

**Abstrak:** Sejak diberlakukannya Ujian Akhir Sekolah Berstandar Nasional (UASBN) pada tahun pelajaran 2007/2008, propinsi diberi kewenangan untuk menyusun soal-soal ujian berdasarkan kisi-kisi yang sama yang dikeluarkan BSNP. Para ahli pengukuran menyatakan bahwa tidak pernah ditemukan dua paket tes dengan butir soal yang berbeda walaupun berdasarkan kisi-kisi yang sama mempunyai tingkat kesukaran yang sama. Oleh karena itu dalam pelaksanaan evaluasi setingkat Ujian Nasional yang menggunakan beberapa paket tes yang berbeda dan mengukur hal yang sama, perlu dilakukan penyetaraan (*equating*). Dengan dilakukannya penyetaraan/*equating* skor atas paket-paket yang digunakan dalam UASBN, maka estimasi kesalahan pengukuran yang sekecil apapun dapat diketahui, dan skor yang diperoleh dapat dibandingkan sehingga peserta tes tidak merasa dirugikan atau diuntungkan karena kebetulan mendapat paket tes yang lebih mudah atau yang lebih sukar. Tujuan Penelitian adalah menentukan konversi nilai mata pelajaran yang diujikan secara nasional antarprovinsi, antarpaket pada jenjang pendidikan Sekolah Dasar, dan mengetahui kemampuan sebenarnya dari peserta UASBN berdasarkan konversi skor yang telah disetarakan. Pada penelitian ini, teknik *equating* yang digunakan adalah *equipercentile equating* dengan menggunakan *software Common Item Program for Equating (CIPE)* versi 2.0.

**Kata kunci:** *UASBN, penyetaraan tes, soal anchor, teori tes klasik, teori respon butir, dan skor konversi*

**Abstract:** Since the enactment of the National Standard School Final Examination (UASBN) in school year 2007/2008, the province is authorized to construct exam questions based on the same lattice issued BSNP. No tests ever found two packages with different grain problem, although based on the same grid have the same difficulty level. Therefore, the evaluation of the National Exam level using several different test packets and measuring the same thing, need to be equated. The effect of equating scores on the packages used in this UN estimates that the slightest error of measurement can be known, and the scores obtained can be compared so that the test participants are not disadvantaged or advantaged because of getting an easier or more difficult test package accidentally. The purpose of this Research is to 1) determine the conversion of the value of the subjects to be tested nationally among province, among packages at UASBN.; 2) know the actual ability of the participants UASBN based on the conversion of the values that have been synchronised. In this research, the technique used is *equipercentile equating* with the use of the software Program for Common Item Equating (CIPE) version 2.0

**Key words:** *UASBN, test equating, anchor item, classical test theory, item response theory, and conversion score*

## Pendahuluan

Ujian Akhir Sekolah Berstandar Nasional (UASBN) diselenggarakan dalam rangka memenuhi amanat Undang-Undang Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional. Tujuan dilaksanakannya UASBN antara lain untuk menilai pencapaian kompetensi lulusan secara nasional (pada mata pelajaran Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam); keterbandingan antar-sekolah, antardaerah, dan antartahun pada mata

pelajaran yang diujikan. Hasil UASBN digunakan sebagai salah satu pertimbangan untuk pemetaan mutu satuan pendidikan, dasar seleksi masuk jenjang pendidikan berikutnya, dan penentuan kelulusan peserta didik dari satuan pendidikan. Nilai-nilai tersebut dapat dibandingkan jika semua peserta tes di setiap provinsi mengerjakan soal-soal (perangkat tes) yang sama. Tetapi dalam pelaksanaannya, pada UASBN diadministrasikan lebih dari satu perangkat tes di setiap provinsi,

dan menggunakan perangkat-perangkat tes berbeda antarprovinsi. Dengan pengadministrasian lebih dari satu perangkat tes, perbedaan nilai antarpeserta tes yang mendapat tes berbeda tidak dapat langsung disimpulkan adanya perbedaan kemampuan antarmereka, karena tingkat kesukaran perangkat yang digunakan akan mempengaruhi perbedaan tersebut.

Untuk menanggulangi masalah tersebut, para ahli *educational measurement* mengembangkan metode statistik yang disebut *equating* (penyetaraan). Metode ini adalah metode ilmiah yang digunakan untuk menyetarakan nilai dari skor mentah satu perangkat ke skor mentah perangkat lainnya, yang pada akhirnya akan diperoleh sebuah tabel konversi nilai. Menurut Suryabrata (1987) bahwa dalam pelaksanaan evaluasi yang menggunakan beberapa perangkat tes yang berbeda dan mengukur hal yang sama, perlu dilakukan penyetaraan dari perangkat-perangkat tes yang digunakan tersebut, karena dengan penyetaraan tersebut dapat dijamin keadilan bagi peserta tes.

Hambleton & Swaminathan (1985) mengatakan bahwa tidak pernah ada soal dari dua perangkat tes dengan butir soal yang berbeda walaupun berdasarkan kisi-kisi yang sama mempunyai tingkat kesukaran yang sama. Penyetaraan skor dapat dijadikan sebagai teknik penyamaan skor yang mampu membedakan peserta didik yang pandai dan peserta didik yang kurang pandai. Penyetaraan skor dimungkinkannya penggunaan perangkat tes yang berbeda terhadap kelompok yang berbeda, sesuai dengan tingkat kemampuannya, sehingga skor yang diperoleh dapat dibandingkan dan peserta tes tidak merasa dirugikan atau diuntungkan karena kebetulan mendapat perangkat tes yang lebih mudah atau yang lebih sukar. Dengan demikian, menjadi suatu keharusan bagi para pengembang tes atau lembaga pengembang tes untuk melakukan penyetaraan terhadap perangkat tes yang digunakan.

Tujuan utama dilakukannya *equating* adalah menegakkan keadilan bagi peserta tes maupun pengguna hasil tes. Asumsi bahwa suatu tes paralel dari segi materi (berasal dari kisi-kisi yang sama) diyakini tidak benar, oleh karena itu perlu suatu proses yang menyetarakan skor perangkat

tes paralel dengan mengeliminasi faktor perbedaan tingkat kesukaran antarperangkat tersebut. Secara proses, *equating* adalah prosedur pemberian skor peserta tes sesuai kemampuannya dengan meniadakan efek perbedaan tingkat kesukaran antarperangkat tes. Hal ini sesuai dengan tuntutan keadilan, jangan sampai peserta didik mendapat nilai jelek karena mengerjakan perangkat tes sukar ataupun peserta didik nilainya baik karena mengerjakan tes yang mudah. Pengguna hasil tes juga menuntut validitas hasil, jangan sampai seseorang mendapatkan hasil tesnya baik hanya karena mengerjakan tes yang lebih mudah padahal kemampuannya rendah

Secara psikometris, *equating* merupakan suatu proses yang bertujuan untuk memperoleh skor konversi yang dapat memperbandingkan hasil beberapa perangkat tes yang paralel. Petersen (1989) mendefinisikan *equating* sebagai prosedur empirik yang dilakukan untuk menghasilkan hubungan antara skor mentah dua perangkat tes yang paralel, yaitu A dan B, sehingga skor perangkat tes A dapat diartikan dalam sudut pandang skor perangkat tes B, sedangkan Angoff (1984) mendefinisikan *equating* sebagai proses untuk mengonversi unit di satu perangkat tes menjadi unit di perangkat tes lain yang paralel.

Proses *equating* adalah hal yang mutlak harus dilakukan dalam menangani pengolahan hasil ujian nasional agar diperoleh pemetaan mutu pendidikan yang akurat dan valid, tanpa distorsi perbedaan tingkat kesulitan walaupun mendapat perangkat tes yang berbeda.

Permasalahan penyetaraan (*equating*) tes dalam UASBN di Indonesia merupakan hal yang perlu dilakukan mengingat mutu pendidikan di wilayah Indonesia yang belum merata, dan keadaan geografis wilayah Indonesia sebagai negara kepulauan yang cukup luas. Dalam pelaksanaan evaluasi setingkat Ujian Nasional yang menggunakan beberapa perangkat tes yang berbeda dan mengukur hal yang sama, penyetaraan (*equating*) skor perlu dilakukan. Dengan dilakukannya penyetaraan skor atas perangkat-perangkat yang digunakan dalam UASBN, maka estimasi kesalahan pengukuran yang sekecil apapun dapat diketahui, dan skor yang diperoleh peserta didik antarperspropinsi dapat diperbandingkan sehingga peserta tes tidak

merasa dirugikan atau diuntungkan karena kebetulan mendapat paket tes yang lebih mudah atau yang lebih sukar.

Penelitian ini dimaksudkan untuk menentukan konversi nilai mata pelajaran yang diujikan secara nasional antarprovinsi, antarpaket pada jenjang pendidikan Sekolah Dasar dan mengetahui kemampuan sebenarnya dari peserta UASBN berdasarkan konversi nilai yang telah disetarakan.

## Kajian Pustaka

### Karakteristik Tes

Tes yang berkualitas baik memiliki karakteristik butir soal dan perangkat tes yang baik pula. Karakteristik soal dapat dilihat dari parameter tingkat kesukaran, daya pembeda, reliabilitas, dan kesalahan pengukuran.

Tingkat kesukaran satu butir soal didefinisikan sebagai proporsi peserta tes yang menjawab benar soal tersebut (teori tes klasik). Angka yang menunjukkan mudah, sedang, atau sukar suatu butir soal disebut indeks tingkat kesukaran dengan nilai antara 0 dan 1. Dalam suatu perangkat tes, tingkat kesukaran didapat dengan cara menghitung rerata seluruh indeks tingkat kesukaran. Daya Pembeda adalah kemampuan satu butir soal untuk membedakan antara peserta tes yang pandai dengan peserta tes yang kurang pandai. Untuk mengetahui daya pembeda butir soal biasanya menggunakan indeks korelasi antara skor butir dengan skor totalnya, seperti teknik *point biserial* dan teknik *biserial*. Daya pembeda suatu perangkat tes adalah rerata dari daya beda butir-butir soal yang membentuk tes tersebut. *Reliabilitas* adalah tingkat kepercayaan dari suatu alat ukur, artinya seberapa jauh pengukuran dilakukan berulang-ulang terhadap sekelompok subyek yang sama memberikan hasil yang sama pula. Ukuran reliabilitas yang baik adalah antara 0,60 sampai dengan 0,85 (Grondlund, 1982). Kesalahan pengukuran (*standart error of measurement*) biasanya disebabkan oleh kesalahan dalam pengambilan sampel. Semakin kecil angka kesalahan semakin baik, sehingga skor yang diperoleh dari perangkat tersebut semakin mendekati skor yang sebenarnya.

### Definisi Penyetaraan (*Equating*)

Weiss (1983) mendefinisikan penyetaraan skor

sebagai suatu prosedur empiris karena data skor diperoleh dari hasil pekerjaan peserta didik yang selanjutnya diperlukan untuk mentransformasi skor. Menurut Hambleton (1991) penyetaraan skor adalah membandingkan skor yang diperoleh dari perangkat tes yang satu (X) dan perangkat tes lainnya (Y) yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tes tersebut. Crocker dan Algina (1986), menyatakan bahwa dua skor hasil pengukuran yang menggunakan instrumen X dan instrumen Y dapat disetarakan skornya jika kedua instrumen mengukur kemampuan atau *trait* yang sama. Menurut Kolen (2004) penyetaraan skor dapat dilakukan jika kelompok peserta tes setara, karena kesetaraan yang ekstrim akan berpengaruh dalam perhitungan.

Berdasarkan pengertian di atas dapat disimpulkan bahwa penyetaraan/*equating* merupakan prosedur secara empiris dalam rangka menyetarakan skor dari perangkat tes yang satu ke perangkat tes lainnya sehingga dapat melakukan perbandingan atau konversi secara langsung hasil-hasil individu yang mengikuti perangkat tes yang berbeda tersebut.

### Tipe-tipe *Equating* / Penyetaraan.

Ada beberapa teknik dan metodologi yang dapat digunakan dalam penyetaraan tes untuk menetapkan transformasi. Secara umum teknik dan metodologi ini dapat dibagi atas dua jenis, yaitu yang metode teori tes klasik (*classical test theory*) dan metode teori modern (*item response theory*).

Menurut Hambleton, Swaminathan, & Roger (1991), pada teori tes klasik dikenal dua metode, yaitu penyetaraan linear (*linear equating*) dan penyetaraan ekuipersentil (*equipercentile equating*). Penyetaraan linear akan menghubungkan skor konversi dengan skor asalnya melalui suatu fungsi linear. Prinsip dasar metode ini adalah distribusi skor pada dua perangkat tes sama dalam hal rerata dan simpangan baku. Angoff (1971) menyatakan bahwa definisi untuk penyetaraan linear adalah skor dua perangkat tes menjadi ekuivalen jika ada hubungan yang setara dengan standar skor deviasinya. Penyetaraan skor yang menggunakan metode linear memungkinkan adanya tingkat kesulitan relatif bervariasi pada

skor di antara beberapa perangkat tes tersebut. Pada penyetaraan linear diperlukan kesamaan distribusi probabilitas antara skor X dan skor Y. Jika skor X dan skor Y memiliki rerata dan simpangan baku yang berbeda, maka distribusi probabilitas yang sama dari kedua skor tersebut dapat digunakan untuk mentransformasi nilai dari satu distribusi probabilitas ke distribusi probabilitas berikutnya.

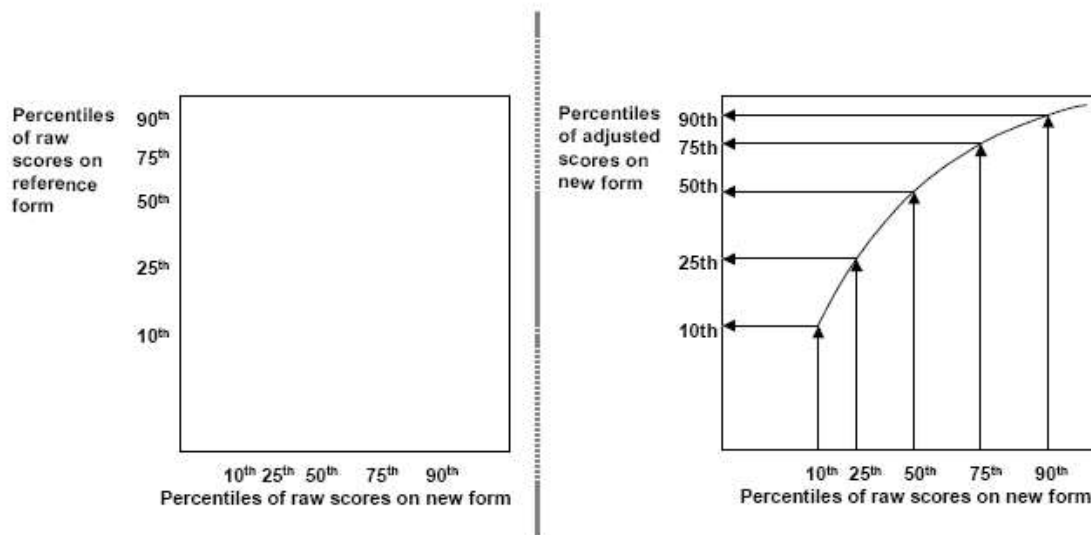
Tipe yang kedua metode penyetaraan ekuipersentil (*equipercentile equating*). Penyetaraan ekuipersentil adalah metode penyetaraan dua paket tes, misal X dan Y, dengan mengasumsikan bahwa kedua paket tersebut mengukur variabel laten yang sama dan nilai persentil rank distribusi skor kedua paket tes hampir sama. Apabila distribusi skor kedua paket tes persis sama, maka hasil *equating* skor di paket X akan persis sama dengan skala skor paket tes Y. Dengan demikian metode penyetaraan ekuipersentil mengkonversi skor berdasarkan jenjang persentil yang sama. Gambar 2.1 berikut adalah ilustrasi prosedur penyetaraan dengan metode ekuipersentil.

Livingstone (1984) menjelaskan prosedur penyetaraan ekuipersentil sebagai berikut: pertama membuat tabulasi ranking persentil untuk distribusi skor pada masing-masing paket yang disetarakan. Kedua, transformasikan skor di paket tes yang baru terhadap paket tes acuan sedemikian rupa sehingga skor yang berpasangan memiliki nilai ranking persentil yang sama.

Prosedur estimasi frekuensi penyetaraan ekuipersentil dikerjakan dengan langsung menyetarakan dua perangkat tes yaitu paket X dan paket Y. Namun demikian, paket X dan paket Y dapat juga disetarakan langsung dengan tes yang ketiga (tes V). Prosedur ini lebih baik digunakan dalam penyetaraan ekuipersentil.

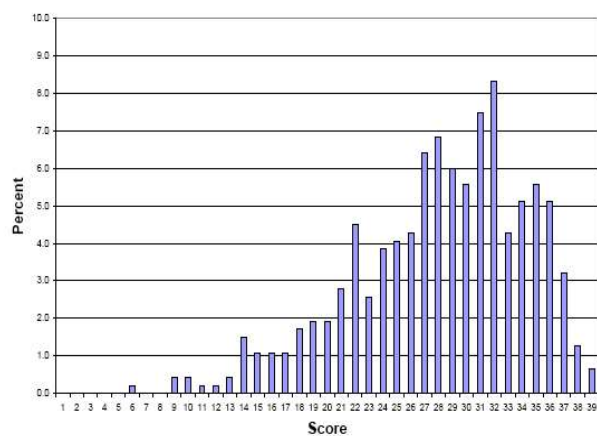
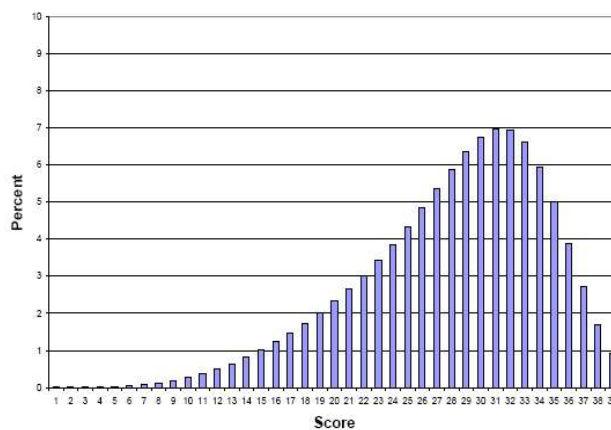
Satu masalah yang sering dihadapi dalam proses penyetaraan/*equating* dengan metode ekuipersentil adalah ketidakteraturan distribusi skor, terutama pada kasus sampel yang kecil. Ketidakteraturan ini menimbulkan masalah karena nilai ranking persentil menjadi tidak stabil saat digeneralisasikan ke populasi. Untuk mengatasi masalah ini perlu dilakukan proses *smoothing* distribusi skor, terutama untuk jumlah sampel yang kecil. *Smoothing* adalah proses pemulusan atau penghalusan ketidakteraturan distribusi skor dengan cara mengganti distribusi tersebut dengan distribusi lain yang memiliki bentuk, lokasi, penyebaran, *skewness*, dan *kurtosis* yang sama, tetapi meminimalisir ketidakteraturan. Gambar 2 menunjukkan ilustrasi proses *smoothing* distribusi skor.

Penerapan teknik pemulusan pada distribusi skor sampel yang mengandung kekeliruan acak dan memiliki bentuk ketidakberaturan (melonjak-lonjak), akan menghasilkan bentuk distribusi skor yang halus. Dengan demikian hasil penyetaraan yang lebih akurat dapat diperoleh dengan cara *smoothing* (pemulusan).



Gambar 1. Percentiles of Raw Score




Sebelum *smoothing*

Setelah *smoothing*

Gambar 2. *Smoothing* (penghalusan) distribusi skor

Metode penyetaraan dengan *item response theory* (IRT) atau teori respon butir, didasarkan asumsi bahwa ada sebuah fungsi matematika yang menggambarkan hubungan antara kemampuan peserta tes dan kemungkinan peserta tes menjawab soal dengan benar. Ada tiga model penyetaraan dengan *item response theory*, yaitu *Rasch Model* (satu parameter logistik), model dua parameter logistik, dan model tiga parameter logistik (Hambleton, 1991). Model Rasch mengasumsikan tingkat kesukaran butir soal sebagai satu-satunya karakteristik yang mempengaruhi kemampuan peserta tes. Pada model dua parameter, kemampuan peserta tes dipengaruhi oleh tingkat kesukaran item dan daya pembeda soal tanpa faktor menebak (*guessing*). Sedangkan pada model tiga parameter, kemampuan peserta tes dipengaruhi oleh tingkat kesukaran, daya pembeda, dan faktor menebak (*guessing*). Prosedur *equipercentile* sama baiknya dengan penyetaraan dengan prosedur IRT.

Dari perspektif yang berbeda, penyetaraan tes dapat dibedakan atas dua tipe, yaitu penyetaraan vertikal dan penyetaraan horizontal. Penyetaraan vertikal didefinisikan sebagai sebuah metode pengukuran nilai pada dua tes yang berbeda tingkatan kelas. Contoh tes Matematika untuk kelas 3 dan kelas 5, pada skala yang sama sehingga nilai Matematika para peserta didik dari dua tingkat yang berbeda tersebut dapat dibandingkan. Kolen (1984) menyatakan bahwa *vertikal equating* memuat perbandingan perkembangan kemampuan peserta didik pada kedua

level pada saat tes dan perkembangan kemampuan dari waktu ke waktu dapat pula dibandingkan. Dengan kata lain pada penyetaraan vertikal dimaksudkan untuk menentukan padanan skor-skor yang diperoleh dari dua kelompok peserta tes dalam tingkat atau jenjang pendidikan yang berbeda, tetapi dikenakan perangkat tes yang sama.

Penyetaraan horizontal didefinisikan sebagai metode skor penempatan peserta tes pada dua tes yang sama di level yang sama, mengukur hal yang sama, dan untuk populasi yang sama sehingga skor peserta tes dapat dibandingkan. Penyetaraan horizontal dimaksudkan untuk menyetarakan dua skor yang masing-masing diperoleh dari dua perangkat tes yang berbeda, tetapi mengukur hal yang sama. Penyetaraan horizontal lebih jelas karena tujuan dari penyetaraan adalah membandingkan dua atau lebih kelompok peserta tes yang memiliki kemampuan yang sama menggunakan dua atau lebih perangkat tes yang berbeda mengukur hal yang sama dan tingkat kesukaran yang sama.

### Desain Penyetaraan/*equating* Tes

Menurut Peterson (1989), penyetaraan tidak sesederhana seperti regresi, karena metode penyetaraan adalah sebuah prosedur empiris yang melibatkan sebuah desain untuk pengumpulan data dan sebuah aturan untuk menetapkan transformasi. Beberapa desain dapat digunakan untuk memperoleh data pada proses penyetaraan/*equating*. Ada tiga desain yang sering

digunakan oleh lembaga testing, yaitu *single group design*, *common item non equivalent*, dan *random group design*.

Pada *single group design*, satu sampel yang sama diuji dua kali dengan paket tes yang berbeda. Misal paket tes tersebut paket X dan paket Y. Paket X diadministrasikan pertama pada semua peserta tes, kemudian disusul dengan paket Y. Cara ini mengakibatkan paket Y lebih sukar soalnya dari paket X karena diujikan belakangan (*fatigue effect*).

Untuk mengurangi *fatigue effect* dapat digunakan *Counterbalancing*, yaitu dengan cara sampel dibagi atas subgrup 1 dan subgrup 2. Paket X diadministrasikan pada peserta pertama, diikuti paket Y. Peserta kedua sebaliknya, mendapat Paket Y, kemudian Paket X. Walaupun *Counterbalancing* dapat digunakan untuk mengurangi *fatigue effect*, *Single group design* jarang digunakan. Salah satu kendala penggunaan *Single group design* dikarenakan desain ini mengharuskan pengadministrasian semua paket tes pada semua peserta, terutama jika ada lebih dari satu perangkat.

Desain yang kedua adalah *common item non equivalent*. Desain ini lebih tepat digunakan untuk ujian nasional, karena ada soal-soal yang sama antarpaket (*forms*) pada UN sebagai *anchor item* dan paket-paket tes tersebut diadministrasikan pada kelompokpeserta yang berbeda. Untuk memperoleh hasil *equating* yang memuaskan perlu jumlah *common item* yang proportional dengan content pada total tes. Crocker dan Algina (1986) mengatakan bahwa jumlah minimal anchor item yang baik adalah 20% dari seluruh item.

Desain yang ketiga adalah *random group design*. Paket-paket tes pada desain ini diadministrasikan secara random/acak kepada semua peserta tes dengan pembagian paket tes secara spiral. Misal ada dua paket yang diujikan yaitu paket X dan paket Y yang dibagikan secara bergantian. Peserta tes pertama akan mendapat paket X, peserta kedua mendapat paket Y, peserta ketiga mendapat paket X, dan seterusnya. Dengan cara ini menjamin perangkat tes akan terdistribusi pada dua grup peserta tes dengan kemampuan yang relatif sama. Salah satu kelebihan dari desain ini adalah setiap peserta tes hanya menempuh satu paket tes sehingga

faktor kelelahan (*fatigue effect*) peserta tes dapat dihilangkan. Selain itu banyak paket yang diadministrasikan/diujikan dalam waktu yang bersamaan.

Dalam penelitian ini desain penyetaran tes yang digunakan adalah desain kedua yaitu *common item non equivalent*. Desain ini lebih tepat digunakan untuk ujian nasional, karena adanya soal-soal yang sama antarpaket (*forms*) pada UN sebagai *anchor item*

### Metodologi Penelitian

Ujian Akhir Sekolah Berstandar Nasional merupakan salah satu penilaian berskala nasional yang menguji populasi. Populasi dalam penelitian ini adalah seluruh peserta UASBN 2009/2010 di seluruh Indonesia. Sampel diambil dari database UASBN 2009/2010. Jumlah sampel UASBN adalah 5.000 siswa dan dipilih secara acak dari seluruh sekolah peserta UASBN di masing-masing provinsi di Indonesia. Sehingga jumlah keseluruhan sampel adalah 33 provinsi x 5000 siswa (165.000 siswa).

Data yang digunakan pada penelitian ini adalah data hasil Ujian Akhir Sekolah Berstandar Nasional (UASBN) tahun 2009/2010 untuk Sekolah Dasar/Madrasah Ibtidaiyah (SD/MI). Adanya keterbatasan jumlah data yang mampu dianalisis oleh software yang digunakan dalam penyetaraan/*equating* ini, maka secara acak dari setiap database respon peserta didik setiap mata pelajaran setiap provinsi dipilih maksimum 5.000 peserta didik. Adapun mata pelajaran yang digunakan adalah seluruh mata pelajaran yang di-UASBN-kan yaitu Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam.

Teknik Analisis yang digunakan adalah analisis *equating* berdasarkan teori tes klasik. Skema tes mengikuti pola "*Non Equivalent Anchor Test*" (NEAT). Karena soal yang diujikan berbentuk tes paralel dan peserta tes berasal dari populasi yang beragam, maka untuk dapat menyetarakan satu paket tes dengan paket tes yang lain, digunakan *anchor soal*, yaitu soal-soal yang sama muncul di beberapa paket tes. Gambaran perbedaan kemampuan peserta didik dilihat dari proporsi peserta didik yang menjawab benar pada soal-soal *anchor* tersebut. Dengan mengacu pada proporsi tersebut, maka perbedaan tingkat

kesukaran pada soal yang unik di setiap paket tes dapat disesuaikan.

Pada penelitian ini, teknik *equating* yang digunakan adalah *equipercentile equating*. Paket tes dari provinsi X yang akan disetarakan dengan provinsi Y, diasumsikan mengukur hal yang sama karena kedua paket tes tersebut berasal dari kisi-kisi yang sama, serta nilai persentil rank distribusi skor kedua paket tes hampir sama. Jika distribusi skor tidak beraturan sehingga nilai ranking persentil menjadi tidak stabil saat digeneralisasikan ke populasi, maka dalam analisis ini dilakukan *smoothing* (penghalusan) distribusi skor. Selanjutnya diperoleh hasil transformasi skor paket tes baru terhadap paket tes acuan. Studi *equating* ini menggunakan *software Common Item Program for Equating* (CIPE) versi 2.0 yang dibuat oleh Kolen (2004).

### Hasil Penelitian dan Pembahasan

Dalam UASBN terdapat tiga mata pelajaran yang diujikan yaitu Bahasa Indonesia, Matematika, dan IPA. Soal UASBN disusun berdasarkan kisi-kisi penulisan soal dari Pusat (Balitbang Diknas). Sebanyak 75% soal ditulis oleh penulis dari masing-masing provinsi dan 25% soal dari Pusat sebagai soal *anchor*. Dari setiap provinsi dipilih secara acak 5.000 peserta didik, sehingga jumlah sampel yang dianalisis di setiap provinsi (N) = 5.000. Sebagai acuan penyetaraan dipilih data dari Provinsi Jawa Barat. Provinsi ini dipilih karena

memiliki jumlah sampel (peserta didik) yang banyak, dan analisis karakteristik soal (dengan menggunakan IRT) menunjukkan hasil yang stabil.

Data yang diolah terdiri atas skor perolehan peserta didik dari total soal yang diujikan, 50 soal untuk mata pelajaran Bahasa Indonesia, dan 40 soal masing-masing untuk mata pelajaran Matematika dan IPA, serta skor perolehan soal *anchor* dengan jumlah soal masing-masing mata pelajaran 10 soal untuk mata pelajaran Matematika dan IPA, dan 13 soal untuk mata pelajaran Bahasa Indonesia. Hasil analisis *univariate* untuk ketiga mata pelajaran tersebut diperoleh *mean anchor* USBN berikut ini.

Tabel 1 menggambarkan *mean anchor* tertinggi untuk **Bahasa Indonesia** diperoleh provinsi DI Yogyakarta (9.79); *mean anchor* terendah diperoleh provinsi Riau (6.36); Jawa Barat sebagai provinsi acuan berada di urutan ke-5 (8.86).

Pada mata pelajaran **Matematika** *mean anchor* tertinggi diperoleh provinsi DI Yogyakarta (7.51); *mean anchor* terendah diperoleh provinsi Sumatera Selatan (3.02); Jawa Barat sebagai provinsi acuan berada di urutan ke-4 (6.77).

Pada mata pelajaran **IPA** *mean anchor* tertinggi diperoleh provinsi DI Yogyakarta (7.18); *mean anchor* terendah diperoleh provinsi Maluku Utara; (5.18)) Jawa Barat sebagai provinsi acuan berada di urutan ke-3 (6.68).

Tabel 1 . Mean Anchor UASBN Tahun Pelajaran 2009/2010

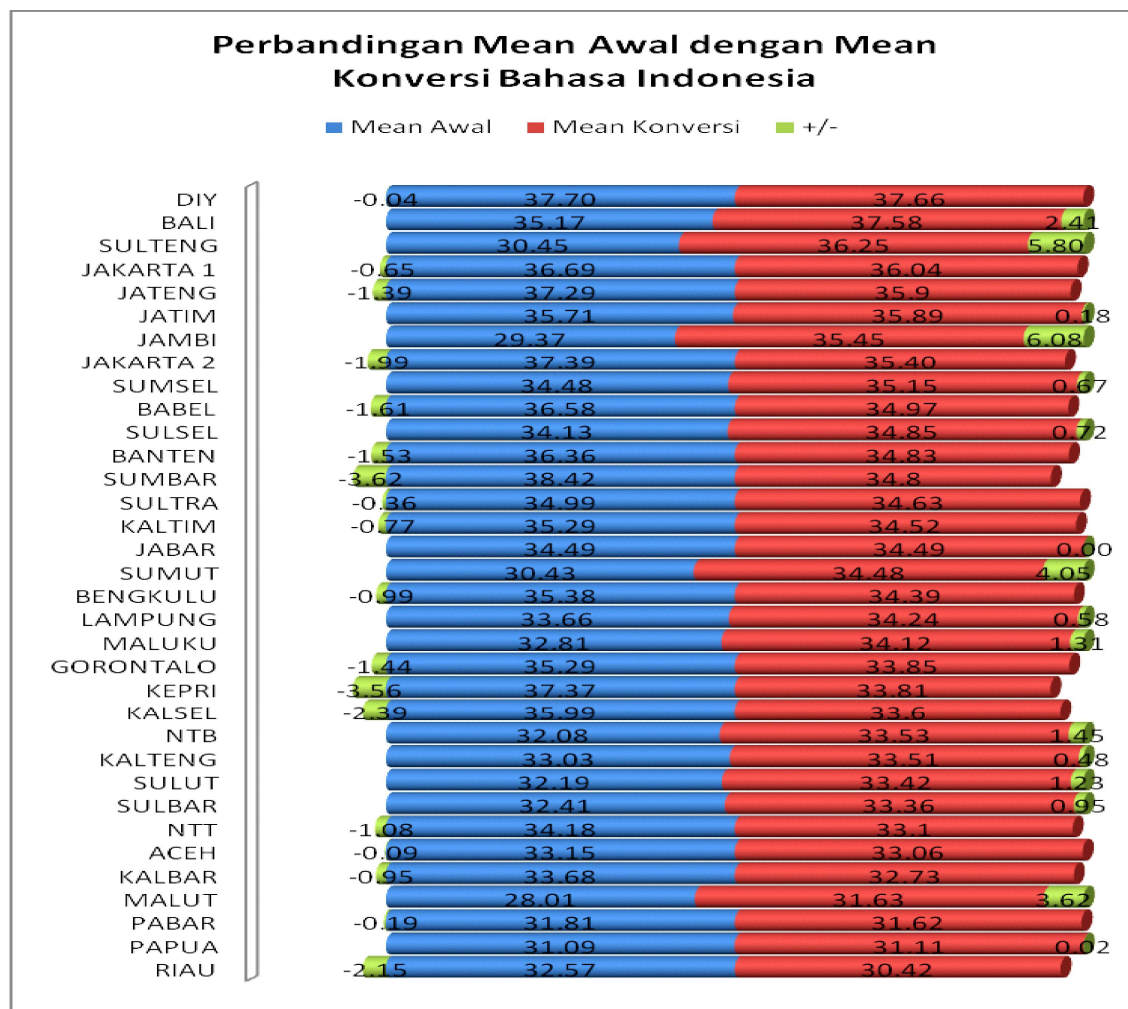
PROVINSI	BIN	MAT	IPA	PROVINSI	BIN	MAT	IPA
DKI Jakarta 01	9.17	6.28	6.27	Sultra	8.19	5.81	6.04
DKI Jakarta 02	8.61	6.27	6.20	Sulteng	8.29	6.00	6.10
Jabar 01	8.86	6.77	6.68	Sulsel	8.78	6.94	6.37
Jateng	8.87	6.75	6.77	Sultra	8.25	6.19	5.91
DIY	9.79	7.51	7.18	Maluku	8.03	5.91	6.31
Jatim	8.84	6.78	6.73	Bali	9.56	7.24	6.22
Aceh	7.69	5.65	6.27	NTB	7.80	5.38	5.55
Sumut	8.24	6.42	6.23	NTT	7.56	5.31	6.01
Sumbar	8.33	6.64	6.45	Kepri	8.13	5.51	6.08
Riau	6.36	5.74	6.03	Bengkulu	8.18	5.84	6.38
Jambi	8.69	6.33	6.50	Maluku Utara	6.60	4.61	5.18
Sumsel	8.28	3.02	6.21	Babel	8.38	5.64	6.44
Lampung	8.10	5.31	5.89	Gorontalo	7.85	5.86	6.11
Kalbar	7.39	4.62	5.67	Banten	8.33	6.20	6.36
Kalteng	7.73	4.57	5.76	Sulbar	7.75	5.56	5.60
Kalsel	7.89	4.97	5.83	Papua Barat	6.96	4.81	5.38
Kaltim	8.24	5.41	5.90	Papua	7.44	4.69	5.46

### Hasil *equating* mata pelajaran Bahasa Indonesia.

Jumlah soal Bahasa Indonesia UASBN SD tahun pelajaran 2009/2010 terdiri dari 50 soal, dengan 13 soal *anchor*. Seluruh soal (50 item) dianalisis untuk mendapatkan hasil konversi penyetaraan (*equate*). Gambar 3 berikut menunjukkan perbandingan mean awal dengan mean konversi untuk setiap provinsi, dengan Jawa Barat sebagai provinsi acuan penyetaraan. DKI Jakarta karena memakai 2 perangkat tes, maka kedua perangkat tersebut juga dianalisis (DKI 1 dan DKI 2).

diperoleh provinsi Riau (30.42) Sementara itu kenaikan mean tertinggi diperoleh provinsi Jambi (sebanyak 6.08 poin) yaitu dari 29.37 menjadi 35.45, dan penurunan mean setelah disetarakan terjadi paling tajam di provinsi Sumatera Barat (sebanyak -3.62) yaitu dari 38.42 menjadi 34.8. Tampak pada gambar 3 di atas terdapat 15 provinsi menunjukkan kenaikan mean yaitu berkisar antara 0.02 sampai dengan 6.08. sisanya mengalami penurunan setelah dikonversi.

Gambar 4 adalah contoh hasil penyetaraan skor peserta didik pada UASBN SD untuk mata

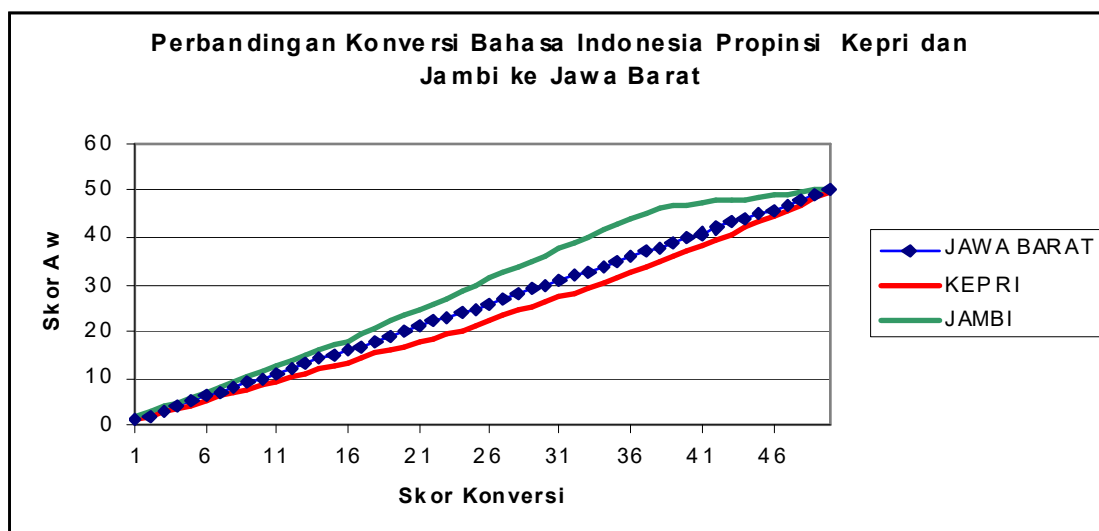


Gambar 3. Perbandingan mean awal dan mean konversi Bahasa Indonesia

Gambar di atas menunjukkan rerata skor dari sebelum disetarakan dan rerata skor setelah disetarakan. Skor mean awal tertinggi adalah provinsi DIY (37.70), dan terendah adalah Maluku Utara (28.01). Setelah disetarakan mean tertinggi tetap diperoleh provinsi DIY (37.66), dan terendah

ujian Bahasa Indonesia di Kepulauan Riau dan Jambi dibandingkan skor peserta didik sebelum dilakukan penyetaraan (provinsi Jawa Barat sebagai provinsi acuan penyetaraan).

Skor konversi adalah skor peserta didik. Garis lurus naik (yang berada di tengah) pada



Gambar 4. Konversi Bahasa Indonesia Propinsi Riau dan Jambi ke Jawa Barat

gambar adalah skor awal Jawa Barat . Garis lurus naik menunjukkan sebagai provinsi acuan Jawa Barat memiliki skor awal dan skor konversi sama. Sementara grafik skor konversi Kepulauan Riau (dalam grafik ditunjukkan sebagai garis di bagian bawah) pada bagian skor awal 0 sampai 1 berimpit dengan grafik garis skor Jawa Barat, namun di atas nilai skor awal 1 grafik garis skor Kepulauan Riau berada di bawah grafik garis skor Jawa Barat. Hal ini menunjukkan bahwa setelah dilakukan penyetaraan skor peserta didik Kepulauan Riau terhadap skor peserta didik Jawa Barat, untuk skor awal 0 sampai 1 akan dikonversikan tetap atau sama yaitu 1 (satu), tetapi untuk skor awal di atas 1 akan dikonversikan lebih rendah. Sehingga dapat disimpulkan berdasarkan hasil UASBN di kedua propinsi tersebut bahwa kemampuan peserta didik di Kepulauan Riau berada di bawah kemampuan peserta didik di Jawa Barat.

Sementara pada grafik garis skor Jambi tampak di atas garis Jawa Barat sebagai propinsi acuan. Skor awal pada grafik di atas menunjukkan di atas skor 1 (satu) posisi skor di atasnya berada di atas garis Jawa Barat. Hal ini menunjukkan bahwa kemampuan peserta didik di Jambi lebih tinggi dibandingkan kemampuan peserta didik di Jawa Barat.

Faktor yang menyebabkan hal ini bisa terjadi, diantaranya adanya soal-soal di provinsi tersebut lebih sukar atau lebih mudah dari soal-soal yang digunakan pada provinsi acuan penyetaraan.

Hasil analisis dalam bentuk tabel konversi untuk skor-skor mata pelajaran **Bahasa Indonesia** di seluruh propinsi yang di-*equite* ke Jawa Barat sebagai propinsi acuan dapat dilihat pada tabel 2 berikut ini. Pada tabel tampak terjadi perubahan skor setelah dikonversi, baik menunjukkan kenaikan maupun penurunan. Terdapat 14 Propinsi yang mengalami penurunan skor setelah dikonversi adalah Bangka Belitung, Banten, Bengkulu, DKI Jakarta 1, DKI Jakarta 2, Gorontalo, Jawa Tengah, Kalimantan Barat, Kalimantan Timur, Kalimantan Selatan, Kepulauan Riau, Papua Barat, Papua, Riau, dan Sumatera Barat.

Penurunan skor tersebut menunjukkan bahwa soal-soal di propinsi tersebut tingkat kesukarannya lebih rendah dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat). Sementara pada 19 propinsi lainnya menunjukkan kenaikan skor, yang menggambarkan bahwa tingkat kesukaran soal-soal di propinsi tersebut lebih tinggi dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat).

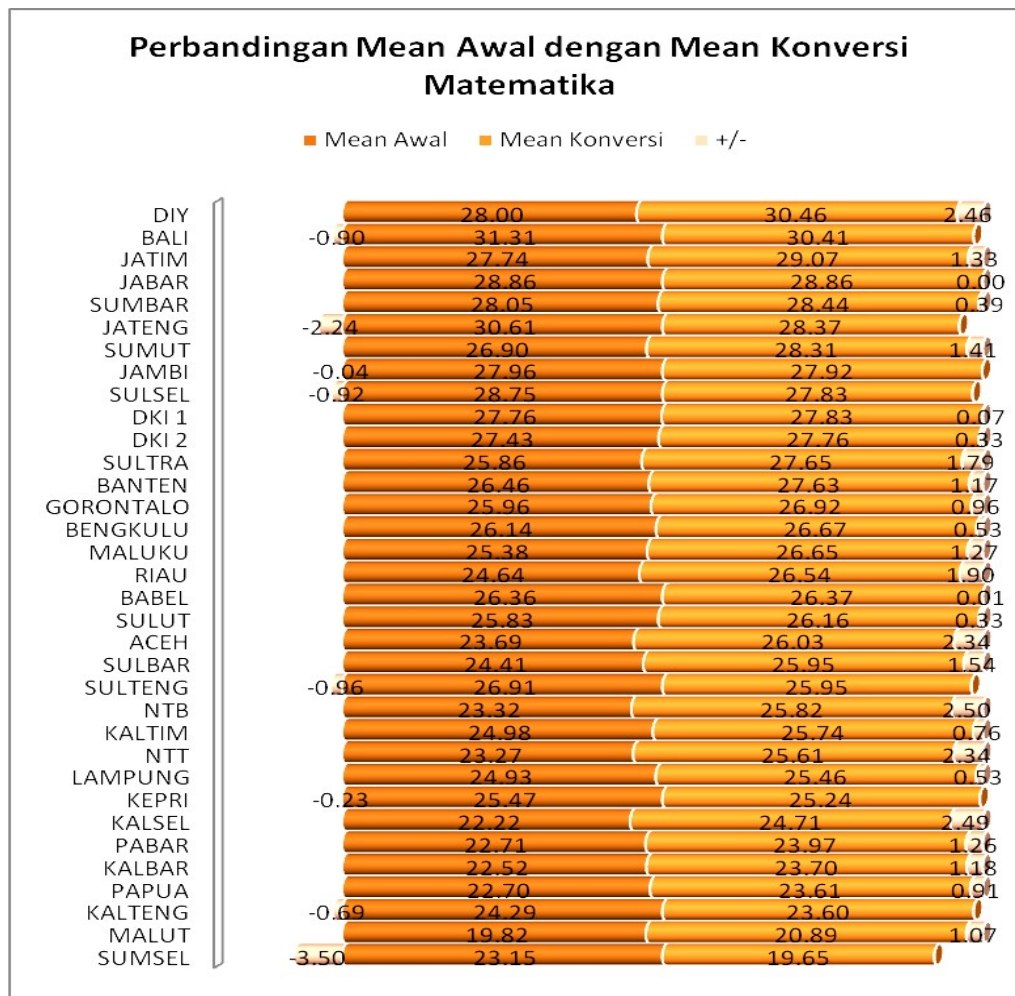
#### Hasil *equating* mata pelajaran Matematika.

Jumlah soal mata pelajaran Matematika UASBN SD tahun pelajaran 2009/2010 terdiri dari 40 soal, dengan 10 soal *anchor*. Seluruh soal (40 item) dianalisis untuk mendapatkan hasil konversi penyetaraan (*equate*).



Tabel 2. Konversi Mata Pelajaran Bahasa Indonesia

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
JABAR	BABEL	BALI	BANTEN	BENGKULU	DKI 1	DKI 2	DIY	GORONTALO	JAMBI	JATENG	JATIM	KALBAR	KALITING	KALITIM	KALSEL	KEPRI	LAMPUNG	MALUKU	MALUT	NAD	NTB	NTT	PABAR	PAPUA	RIAU	SULBAR
1	1.04	1.18	1.04	1.1	1.26	1.14	1.12	1.01	1.42	1.15	1.27	1.11	1.15	1.21	1.00	1.00	1.2	1.2	1.55	1.1	1.3	1.14	1.85	1.17	1.04	1.41
2	1.89	2.14	1.91	2	2.27	2.06	2.04	1.86	2.53	2.09	2.28	2.01	2.09	2.19	1.86	1.83	2.11	2.16	2.75	1.99	2.34	2.07	3.25	3.11	1.90	2.52
3	2.75	3.09	2.77	2.89	3.28	2.99	2.95	2.7	3.65	3.03	3.29	2.92	3.02	3.16	2.7	2.66	3.13	3.12	3.85	2.89	3.38	3	4.65	4.45	2.76	3.62
4	3.61	4.05	3.63	3.79	4.28	3.91	3.87	3.54	4.76	3.96	4.3	3.82	3.96	4.14	3.4	3.49	4.1	4.09	5.16	3.79	4.41	3.93	6.04	5.8	3.62	4.73
5	4.46	5	4.5	4.69	5.29	4.84	4.78	4.38	5.87	4.9	5.31	4.73	4.89	5.11	4.38	4.32	5.07	5.05	6.36	4.69	5.45	4.86	7.44	7.14	4.48	5.83
6	5.32	5.95	5.36	5.59	6.3	5.76	5.7	5.22	6.99	5.84	6.32	5.63	5.83	6.09	5.23	5.15	6.04	6.02	7.56	5.58	6.49	5.79	8.84	8.48	5.34	6.94
7	6.18	6.91	6.22	6.49	7.31	6.69	6.61	6.07	8.1	6.77	7.33	6.53	6.76	7.06	6.07	5.98	7	6.98	8.76	6.48	7.52	6.72	10.24	9.83	6.20	8.05
8	7.04	7.86	7.08	7.38	8.31	7.61	7.53	6.91	9.21	7.71	8.34	7.44	7.69	8.04	6.91	6.81	7.91	7.94	9.96	7.38	8.56	7.65	11.64	11.17	7.06	9.15
9	7.89	8.82	7.95	8.28	9.32	8.54	8.44	7.75	10.33	8.65	9.35	8.34	8.63	9.01	7.75	7.64	8.94	8.91	11.16	8.28	9.6	8.58	13.04	12.51	7.92	10.28
10	8.75	9.77	8.81	9.18	10.33	9.47	9.36	8.59	11.44	9.58	10.36	9.25	9.56	9.99	8.9	8.47	9.9	9.87	12.38	9.17	10.63	9.51	14.44	13.86	8.78	11.38
11	9.61	10.73	9.67	10.08	11.34	10.39	10.27	9.43	12.55	10.92	11.37	10.15	10.5	10.96	9.4	9.30	10.87	10.83	13.43	10.07	11.07	10.44	13.44	14.06	9.64	12.47
12	10.46	11.68	10.53	10.98	12.34	11.32	11.19	10.28	13.67	11.46	12.39	11.06	11.43	11.94	10.28	10.13	11.88	11.8	14.5	10.97	12.71	11.37	16.44	15.87	10.50	13.58
13	11.32	12.64	11.4	11.87	13.35	12.24	12.1	11.12	14.78	12.39	13.4	11.96	12.37	12.92	11.12	10.36	12.4	12.76	15.56	11.87	13.74	12.3	17.45	16.87	11.36	14.68
14	12.18	13.59	12.26	12.77	14.36	13.17	13.02	11.96	15.89	13.33	14.41	12.87	13.3	13.89	11.96	11.79	13.77	13.73	16.63	12.76	14.76	13.23	18.45	17.87	12.32	15.69
15	13.06	14.55	13.12	13.67	15.37	14.09	13.94	12.8	17.01	14.27	15.42	13.77	14.24	14.87	12.81	12.62	14.74	14.69	17.7	13.66	15.77	14.14	19.36	18.88	13.08	16.7
16	13.99	15.5	13.99	14.57	16.38	15.02	14.85	13.64	18.12	15.2	16.43	14.68	15.17	15.84	13.65	13.45	15.1	15.65	18.76	14.63	16.79	15.06	20.12	19.68	14.02	17.71
17	14.75	16.45	14.85	15.47	17.38	15.94	15.77	14.49	19.35	16.14	17.41	15.58	16.11	16.82	14.49	14.28	16.67	16.88	19.78	15.6	17.81	15.97	20.87	20.46	14.95	18.72
18	15.61	17.41	15.71	16.38	18.39	16.87	16.68	15.33	20.67	17.08	18.39	16.56	17.04	17.78	15.36	15.11	17.67	17.7	20.79	16.57	18.82	16.89	21.63	21.24	15.89	19.69
19	16.46	18.36	16.67	17.3	19.39	17.79	17.6	16.3	21.99	18.01	19.37	17.54	17.98	18.75	16.33	15.94	18.64	18.73	21.8	17.54	19.85	17.8	22.38	22.02	16.82	20.65
20	17.38	19.31	17.64	18.22	20.39	18.71	18.51	17.27	23.31	18.96	20.35	18.52	18.91	19.72	17.11	16.77	19.67	19.8	22.8	18.51	20.89	18.72	23.14	22.81	17.15	21.6
21	18.3	20.44	18.6	19.14	21.36	19.64	19.43	18.24	24.62	19.93	21.33	19.53	20.11	20.7	17.98	17.64	20.71	20.89	23.81	19.5	21.94	19.66	23.9	23.59	18.69	22.55
22	19.21	21.56	19.61	20.13	22.34	20.57	20.48	19.23	25.94	20.9	22.31	20.59	21.3	21.67	18.86	18.52	21.77	21.99	24.81	20.6	22.98	20.63	24.66	24.37	19.66	23.51
23	20.25	22.81	20.67	21.15	23.31	21.50	21.53	20.32	27.26	21.88	23.29	21.65	22.48	22.65	19.79	19.39	22.8	23.09	25.82	21.71	24.03	21.6	25.42	25.19	20.67	24.46
24	21.3	24.01	21.73	22.17	24.29	22.44	22.38	21.41	28.58	22.86	24.27	22.71	23.65	23.62	20.82	20.37	23.9	24.19	26.82	22.82	23.82	22.58	26.19	25.93	21.68	25.41
25	22.36	25.32	22.78	23.2	25.26	23.37	23.73	22.5	29.9	23.83	25.26	23.77	24.8	24.6	21.85	21.36	24.97	25.3	27.83	23.93	26.14	23.56	26.96	26.72	22.69	26.37
26	23.41	26.58	23.84	24.24	26.23	24.31	24.88	23.6	31.21	24.81	26.24	24.83	25.94	25.57	22.9	22.36	26.06	26.41	28.84	25.04	27.21	24.54	27.14	27.51	23.69	27.33
27	24.48	27.83	24.9	25.28	27.21	25.25	26.03	24.69	32.52	25.79	27.23	25.89	27.06	26.55	23.95	23.36	27.14	27.53	29.84	26.17	28.28	25.53	28.52	28.3	24.70	28.28
28	25.55	29.08	25.95	26.34	28.16	26.19	27.19	25.78	33.82	26.77	28.21	26.95	28.18	27.52	25.91	24.36	28.23	28.65	30.85	27.29	29.35	26.53	29.32	29.09	25.71	29.24
29	26.62	30.31	27.01	27.41	29.15	27.13	28.34	26.87	35.12	27.75	29.2	28	28.28	28.5	26.89	25.37	29.33	29.79	31.65	28.43	30.43	27.53	30.13	29.89	26.13	30.2
30	27.71	31.55	28.06	28.49	30.13	28.08	29.49	27.96	36.41	28.74	30.19	29.05	30.37	29.47	27.17	26.39	30.46	30.93	32.98	29.58	31.51	28.54	30.95	30.69	27.14	31.16
31	28.8	32.71	29.12	29.58	31.1	29.03	30.64	29.06	37.7	29.72	31.18	30.1	31.44	30.45	28.26	27.42	31.53	32.08	33.96	30.73	32.69	29.56	31.79	31.49	28.16	32.12
32	29.89	33.91	30.18	30.69	32.08	29.99	31.78	30.15	38.97	30.71	32.17	31.14	32.51	31.42	29.36	28.46	32.62	33.23	34.87	31.9	33.66	30.6	32.64	32.31	29.17	33.08
33	30.99	35.15	31.23	31.81	33.05	30.96	32.91	31.24	40.23	31.7	33.17	32.17	33.57	31.4	30.47	29.52	33.71	34.38	35.88	33.07	34.72	31.64	33.5	33.13	30.19	34.04
34	32.09	36.31	32.29	32.94	34.02	31.94	34.03	32.33	41.49	32.69	34.17	33.19	34.63	33.37	31.58	30.59	34.79	35.63	36.88	34.25	35.77	32.69	34.38	33.95	31.82	35.01
35	33.2	37.44	33.35	34.07	35	32.92	35.13	33.43	42.74	33.68	35.16	34.2	35.68	34.35	32.7	31.68	35.87	36.67	37.89	35.44	36.8	33.76	35.27	34.79	32.85	35.97
36	34.3	38.51	34.4	35.2	35.97	33.92	36.22	34.52	43.99	34.67	36.17	35.2	36.74	35.32	33.83	32.78	36.92	37.8	38.9	36.64	37.81	34.83	36.18	35.64	33.88	36.93
37	35.41	39.63	35.46	36.33	36.94	34.93	37.29	35.61	45.23	35.67	37.17	36.19	37.8	36.3	34.46	33.40	37.96	38.92	39.9	37.84	38.79	35.92	37.1	36.5	34.81	37.84
38	36.51	40.69	36.52	37.44	37.92	35.94	38.34	36.7	46.25	36.67	38.17	37.17	38.85	37.27	36.1	35.04	38.99	40.02	40.91	39.04	38.75	37.02	38.03	37.37	35.95	38.64
39	37.61	41.71	37.57	38.54	38.89	36.98	39.37	37.79	46.59	37.68	39.18	38.14	39.9	38.25	37.24	36.19	40	41.1	41.91	40.24	40.69	38.13	38.97	38.26	37.00	39.78
40	38.7	42.71	38.63	39.62	39.86	38.02	40.39	38.89	46.93	38.68	40.19	39.1	40.93	39.22	38.37	37.35	40.99	42.17	42.92	41.44	41.61	39.24	38.93	39.16	38.04	40.74
41	39.79	43.74	39.68	40.68	40.84	39.07	41.39	39.98	47.27	39.69	41.2	40.04	41.95	40.2	39.5	38.52	42.99	43.22	43.93	42.64	42.5	40.37	40.89	40.07	39.09	41.68
42	40.88	44.73	40.74	41.72	41.81	40.13	42.38	41.07	47.61	40.7	42.21	40.98	42.95	41.17	40.82	39.69	42.96	44.27	44.78	43.84	43.38	41.5	41.85	40.99	40.15	42.63
43	41.97	45.71	41.8	42.75	42.79	41.19	43.37	42.16	47.95	41.71	43.22	41.92	43.93	42.15	41.72	40.87	43.92	45.32	45.64	45.03	44.25	42.63	42.82	41.92	41.20	43.57
44	43.05	46.51	42.85	43.77	43.76	42.26	44.35	43.26	48.29	42.72	44.23	42.85	44.89	43.12	42.12	42.06	44.89	46.12	46.29	45.94	45.12	42.63	43.78	42.85	42.16	44.51
45	44.14	47.12	43.91	44.79	44.73	43.32	45.33	44.35	49.63	43.73	45.24	43.77	46.85	44.1	43.91	43.25	44.99	46.79	46.93	46.64	45.99	44.89	44.75	43.8	43.45	45.48
46	45.22	47.7																								



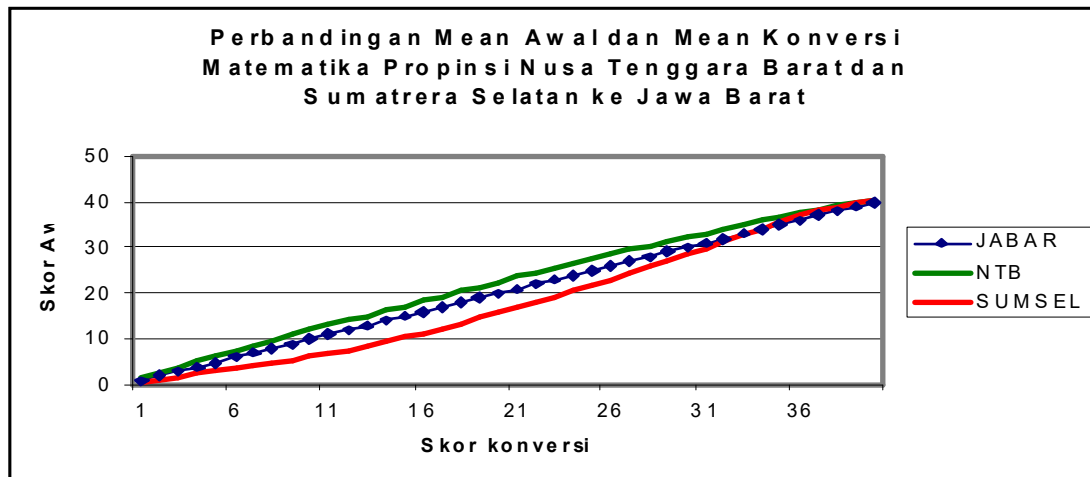
Gambar 5. Perbandingan Mean Awal dengan Mean Konversi Matematika

Gambar 5 di bawah ini menunjukkan rerata (*mean*) skor sebelum disetarakan dan rerata skor setelah disetarakan. Skor mean awal tertinggi adalah provinsi DIY (28.00), dan terendah adalah Maluku Utara (19.82). Setelah disetarakan mean tertinggi tetap diperoleh provinsi DIY (30.46), dan terendah diperoleh provinsi Sumatera Selatan (19.65) Sementara itu kenaikan mean tertinggi diperoleh provinsi Nusa Tenggara Barat (NTB) (sebanyak 2.50 poin) yaitu dari 23.32 menjadi 25.82, dan penurunan mean setelah disetarakan terjadi paling tajam di provinsi Sumatera Selatan (sebanyak -3.50) yaitu dari 23.15 menjadi 19.65. Tampak pada gambar 5 di atas terdapat 25 provinsi menunjukkan kenaikan mean konversi yaitu berkisar antara 0.01 sampai dengan 2.50. sisanya mengalami penurunan setelah dikonversi.

Gambar 6 adalah contoh hasil *penyetaraan* (*equating*) skor peserta didik pada UASBN SD

untuk mata ujian Matematika di Nusa Tenggara Barat dan Sumatera Selatan dibandingkan skor peserta didik sebelum dilakukan penyetaraan (provinsi Jawa Barat sebagai provinsi acuan penyetaraan).

Garis lurus naik (yang berada di tengah) pada gambar 6 adalah skor awal Jawa Barat. Garis lurus naik menunjukkan sebagai provinsi acuan Jawa Barat memiliki skor awal dan skor konversi sama. Sementara grafik skor konversi Sumatera Selatan (dalam grafik ditunjukkan sebagai garis di bagian bawah) pada bagian skor awal 0 sampai dengan skor 34 berada di bawah garis skor provinsi acuan (Jawa Barat). Namun di atas nilai skor awal 34 tampak grafik garis skor Sumatera Selatan naik berada di atas grafik garis skor Jawa Barat. Hal ini menunjukkan bahwa setelah dilakukan penyetaraan skor peserta didik Sumatera Selatan terhadap skor peserta didik Jawa Barat, untuk



Gambar 6. Perbandingan Mean Awal dan Mean Konversi Matematika di Nusa Tenggara Barat dan Sumatera Selatan ke Jawa Barat

skor awal 0 sampai 35 akan dikonversikan lebih rendah, tetapi untuk skor awal di atas 34 akan dikonversikan lebih tinggi. Sehingga dapat disimpulkan berdasarkan hasil UASBN kemampuan peserta didik di Sumatera Selatan sampai dengan skor 34 berada di bawah kemampuan peserta didik di Jawa Barat.

Sementara garis skor Nusa Tenggara Barat (NTB) berada di atas garis Jawa Barat sebagai propinsi acuan. Skor awal pada grafik menunjukkan mulai skor 1 (satu) menunjukkan posisi di atas grafik garis Jawa Barat. Hal ini menunjukkan bahwa kemampuan peserta didik di NTB lebih tinggi dibandingkan kemampuan peserta didik di Jawa Barat.

Hasil analisis dalam bentuk tabel konversi mata pelajaran **Matematika** untuk skor-skor di seluruh propinsi yang di *equite* ke Jawa Barat sebagai propinsi acuan dapat dilihat pada Tabel 3 berikut ini.

Pada Tabel 3 tampak terjadi perubahan skor setelah dikonversi, baik menunjukkan kenaikan maupun penurunan. Terdapat 11 propinsi yang mengalami penurunan skor setelah dikonversi yaitu Bangka Belitung, Bali, DKI Jakarta 1, DKI Jakarta 2, Jawa Tengah, Kalimantan Barat, Kalimantan Timur, Kalimantan Selatan, Kepulauan Riau, Papua, Riau, dan Sulawesi Utara.

Penurunan skor tersebut menunjukkan bahwa soal-soal di propinsi tersebut tingkat kesukarannya lebih rendah dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat). Sementara pada 22 propinsi lainnya

menunjukkan kenaikan skor, yang menggambarkan bahwa tingkat kesukaran soal-soal di propinsi tersebut lebih tinggi dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat).

#### Hasil *equating* mata pelajaran Ilmu Pengetahuan Alam (IPA)

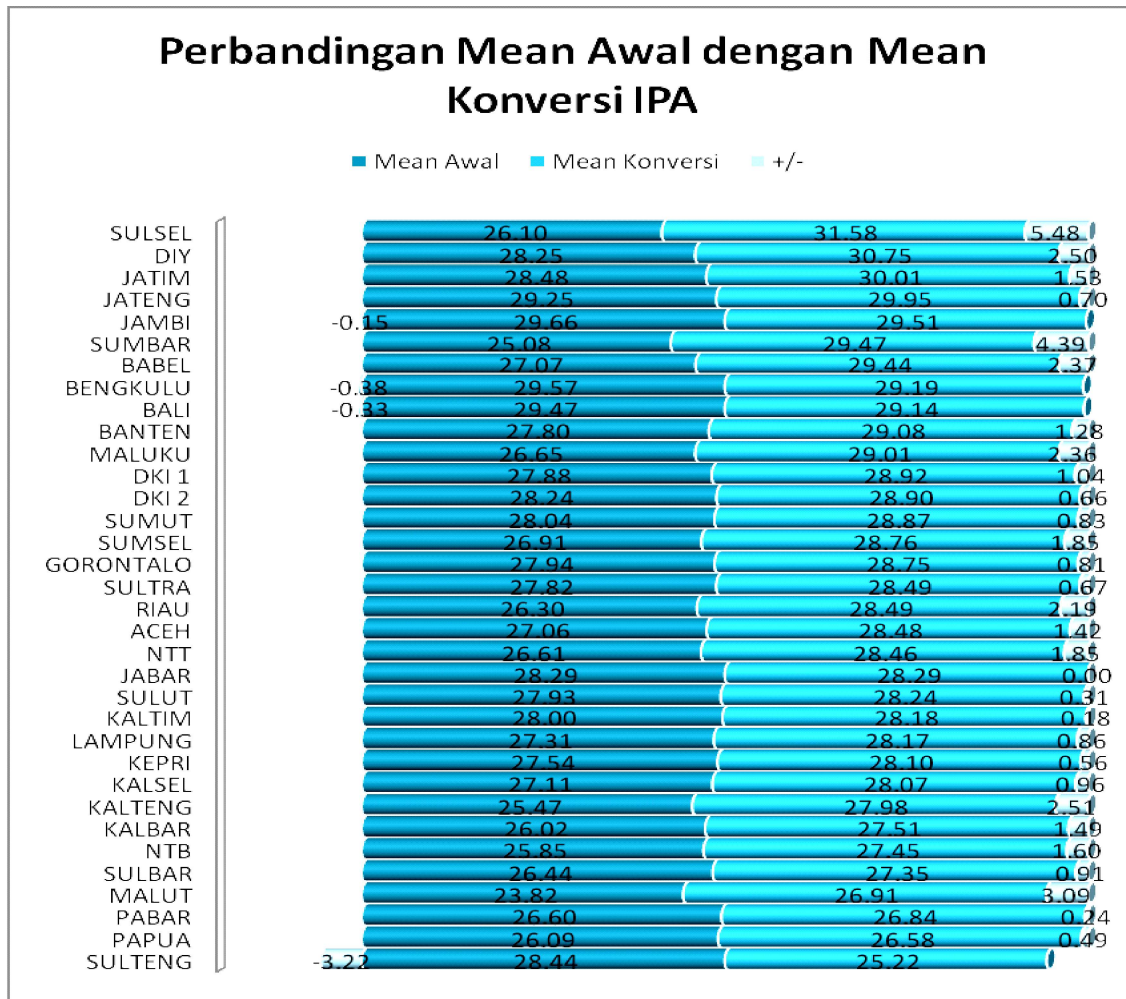
Jumlah soal mata pelajaran IPA UASBN SD tahun pelajaran 2009/2010 terdiri dari 40 soal, dengan 10 soal *anchor*. Seluruh soal (40 item) dianalisis untuk mendapatkan hasil konversi penyetaraan (*equate*). Gambar 7 berikut menunjukkan perbandingan mean awal dengan mean konversi untuk setiap provinsi, dengan Jawa Barat sebagai provinsi acuan penyetaraan.

Gambar 7 menunjukkan rerata skor dari sebelum disetarakan dan rerata skor setelah disetarakan. Skor mean awal tertinggi adalah provinsi Bengkulu (29.57), dan terendah adalah Maluku Utara (23.82). Setelah disetarakan mean tertinggi tetap diperoleh provinsi Sulawesi Selatan (31.58), dan terendah diperoleh provinsi Sulawesi Tengah (25.22). Sementara itu kenaikan mean tertinggi diperoleh provinsi Sulawesi Selatan (sebanyak 5.48 poin) yaitu dari 26.10 menjadi 31.58, dan penurunan mean setelah disetarakan terjadi paling tajam di provinsi Sulawesi Tengah (sebanyak -3.22) yaitu dari 28.44 menjadi 25.22. Tampak pada gambar 7 di atas terdapat 26 provinsi menunjukkan kenaikan mean konversi yaitu berkisar antara 0.18 sampai dengan 5.48. sisanya mengalami penurunan setelah dikonversi. Gambar 8 berikut adalah contoh hasil *penyetaraan*



Tabel 3. Tabel Konversi Mata Pelajaran Matematika

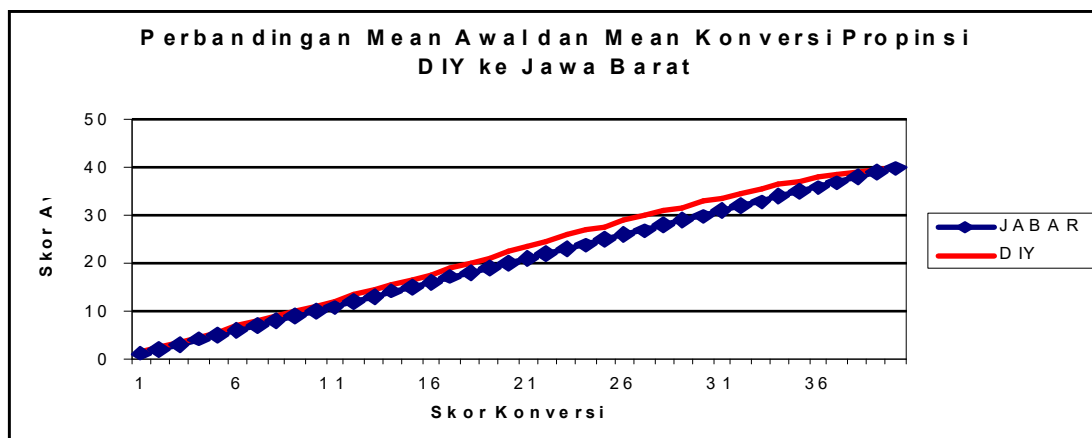
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
JABAR	BABE	BALI	BANTEN	BENGKULU	DIY01	DKI01	DKI02	GORONTALO	JAMBI	JATENG	JATIM	KALBAR	KALSEL	KALTENG	MALTIM	KEPRI	LAMPUNG	MALUKU	MALUT	ACEH	NTB	NTT	PABAR	PAPUA	RIAU	SULBAR
1	1.18	0.97	1.31	1.29	1.62	1.42	1.45	1.44	1.09	1.13	1.72	1.16	1.76	0.85	1.36	1.21	1.12	1.3	1.59	1.64	1.52	1.55	1.72	1.46	1.61	1.45
2	2.13	1.78	2.36	2.32	2.87	2.54	2.58	2.57	1.99	2.05	3.03	1.09	3.1	1.58	2.43	1.18	1.03	2.33	2.81	2.9	2.71	2.75	3.03	2.39	2.85	2.58
3	3.08	2.59	3.4	3.35	4.11	3.65	3.71	3.7	2.88	2.97	4.34	3.03	4.44	2.31	3.51	1.16	1.95	3.36	4.04	4.16	3.89	3.95	4.35	3.73	4.09	3.71
4	4.03	3.41	4.44	4.38	5.36	4.77	4.84	4.82	3.77	3.89	5.65	3.97	5.78	3.04	4.58	4.13	1.86	4.39	5.26	5.42	5.07	5.15	5.66	4.87	5.33	4.84
5	4.98	4.22	5.48	5.4	6.61	5.88	5.97	5.95	4.67	4.81	6.96	4.9	7.12	3.77	5.66	5.11	4.77	5.43	6.49	6.68	6.25	6.35	6.97	6	6.57	5.97
6	5.93	5.03	6.52	6.43	7.85	6.99	7.1	7.08	5.56	5.74	8.27	5.84	8.46	4.5	6.73	5.08	5.68	6.46	7.71	7.94	7.44	7.55	8.28	7.14	7.81	7.1
7	6.88	5.84	7.57	7.46	9.1	8.11	8.24	8.21	6.46	6.66	9.58	6.78	9.8	5.23	7.8	7.05	6.6	7.49	8.94	9.2	8.62	8.74	9.36	8.38	9.05	8.23
8	7.84	6.66	8.61	8.49	10.34	9.22	9.37	9.33	7.35	7.58	10.89	7.72	10.99	5.96	8.88	8.03	7.51	8.52	10.15	10.46	9.8	9.83	10.44	9.34	10.3	9.36
9	8.79	7.47	9.65	9.52	11.6	10.34	10.5	10.46	8.25	8.5	11.9	8.65	12.11	6.7	9.97	9	8.42	9.55	11.36	11.57	10.91	10.92	11.44	10.4	11.44	10.47
10	9.74	8.28	10.69	10.55	12.96	11.46	11.58	11.46	9.14	9.42	12.89	9.59	13.23	7.43	11.07	9.98	9.34	10.58	12.6	12.65	11.98	11.94	12.36	11.45	12.57	11.58
11	10.69	9.1	11.77	11.59	14.24	12.64	12.67	12.42	10.03	10.34	13.87	10.69	14.34	8.16	12.22	11.03	10.25	11.66	13.83	13.73	13.05	12.96	13.28	12.49	13.7	12.7
12	11.73	9.91	12.86	12.66	15.5	13.8	13.76	13.38	11	11.27	14.85	11.92	15.43	8.89	13.36	12.18	11.29	12.78	15.05	14.81	14.12	13.98	14.2	13.52	14.81	13.81
13	12.82	10.72	13.95	13.73	16.74	14.94	14.84	14.35	12.06	12.29	15.82	13.32	16.5	9.62	14.5	13.33	12.48	13.9	16.25	15.88	15.18	15.01	15.12	14.56	15.91	14.91
14	13.91	11.58	15.05	14.8	17.94	16.06	15.9	15.31	13.15	13.29	16.78	14.71	17.54	10.35	15.62	14.48	13.68	15.01	17.44	16.94	16.24	16.03	16.05	15.58	16.99	16.01
15	14.99	12.55	16.14	15.86	19.11	17.13	16.94	16.77	14.23	14.3	17.74	16.06	18.55	11.08	16.73	15.62	14.86	16.12	18.61	17.98	17.29	17.05	16.96	16.61	18.03	17.11
16	16.07	13.64	17.22	16.9	20.24	18.17	17.95	17.24	15.31	15.29	18.69	17.36	19.52	12.11	17.81	16.74	16.04	17.23	19.75	19.02	18.34	18.07	17.88	17.62	19.05	18.17
17	17.14	14.72	18.31	17.94	21.33	19.15	18.92	18.2	16.4	16.28	19.62	18.6	20.47	14.09	18.85	17.85	17.2	18.34	20.86	20.03	19.38	19.09	18.8	18.63	20.03	19.21
18	18.2	15.8	19.38	18.96	22.37	20.09	19.87	19.16	17.48	17.26	20.53	19.78	21.38	15.94	19.86	18.94	18.34	19.43	21.95	21.03	20.42	20.11	19.73	19.63	20.98	20.22
19	19.24	16.88	20.45	19.97	23.37	20.98	20.78	20.13	18.56	18.22	21.44	20.9	22.26	17.62	20.83	20	19.46	20.52	23.02	22.02	21.45	21.14	20.65	20.61	21.9	21.21
20	20.26	17.96	21.51	20.97	24.34	21.83	21.66	21.09	19.64	19.16	22.33	21.96	23.11	19.15	21.76	21.03	20.56	21.59	24.08	22.99	22.47	22.16	21.58	21.58	22.8	22.17
21	21.27	19.05	22.55	21.95	25.26	22.65	22.52	22.05	20.71	20.09	23.2	22.97	23.94	20.54	22.66	22.03	21.63	22.66	25.13	23.94	23.48	23.18	23.51	23.54	23.68	23.11
22	22.25	20.13	23.59	22.92	26.16	23.44	23.35	23.02	21.78	21	24.07	23.94	24.76	21.8	23.54	23.01	22.69	23.71	26.19	24.89	24.49	24.2	23.46	23.48	24.55	24.04
23	23.23	21.21	24.6	23.88	27.02	24.2	24.17	23.98	22.84	21.9	24.92	24.87	25.56	22.95	24.39	23.95	23.73	24.74	27.25	25.83	25.49	25.22	24.41	24.42	25.4	24.95
24	24.19	22.29	25.6	24.83	27.86	24.94	24.97	24.94	23.89	22.79	25.78	25.76	26.35	24.01	25.21	24.87	24.75	25.77	28.32	26.76	26.49	26.24	25.37	25.34	26.24	25.85
25	25.13	23.37	26.58	25.78	28.68	25.67	25.77	25.91	24.94	23.67	26.63	26.64	27.13	25	26.03	25.78	25.75	26.78	29.41	27.69	27.48	27.27	26.34	26.25	27.08	26.75
26	26.07	24.46	27.53	26.71	29.48	26.4	26.57	26.87	25.99	24.54	27.49	27.48	27.92	25.94	26.83	26.66	26.75	27.77	30.5	28.61	28.45	28.29	27.32	27.15	27.91	27.64
27	27	25.54	28.47	27.64	30.27	27.12	27.37	27.83	27.03	25.41	28.35	28.31	28.71	26.84	27.62	27.55	27.73	28.76	31.6	29.54	29.42	29.31	28.31	28.05	28.73	28.54
28	27.98	26.62	29.39	28.56	31.06	27.86	28.18	28.8	28.06	26.29	29.21	29.13	29.5	27.74	28.4	28.42	28.7	29.73	32.71	30.46	30.38	30.33	29.31	28.95	29.55	29.44
29	28.85	27.7	30.29	29.49	31.86	28.6	29	29.76	29.1	27.17	30.09	29.93	30.3	28.66	29.18	29.3	28.67	30.69	33.8	31.37	31.32	31.35	30.31	29.84	30.35	30.35
30	29.78	28.79	31.18	30.41	32.65	29.37	29.83	30.72	30.13	28.08	30.98	30.71	31.1	29.61	29.97	30.17	30.62	31.64	34.89	32.28	32.24	32.37	31.32	30.73	31.16	31.27
31	30.7	29.87	32.05	31.35	33.46	30.15	30.68	31.69	31.16	29	31.88	31.49	31.91	30.6	30.76	31.06	31.58	32.58	35.97	33.19	33.15	33.4	32.31	31.62	31.95	32.2
32	31.68	30.95	32.92	32.29	34.26	30.96	31.55	32.65	32.2	29.94	32.79	32.26	32.74	31.65	31.55	31.95	32.54	33.52	37.04	34.1	34.05	34.42	33.31	32.51	32.75	33.13
33	32.58	32.03	33.79	33.26	35.07	31.81	32.44	33.61	33.24	30.92	33.71	33.03	33.58	32.74	33.37	33.86	33.5	34.44	38.11	34.99	34.93	35.44	34.3	33.4	33.55	34.08
34	33.53	33.11	34.66	34.25	35.89	32.69	33.36	34.58	34.28	31.93	34.64	33.8	34.44	33.87	33.21	33.79	34.46	35.36	38.78	35.89	35.81	36.46	35.28	34.3	34.35	35.02
35	34.49	34.2	35.53	35.27	36.72	33.62	34.31	35.54	35.33	32.98	35.57	34.58	35.31	35.02	34.08	34.75	35.43	36.27	39.41	36.77	36.68	37.48	36.26	35.19	35.18	35.98
36	35.47	35.28	36.41	36.31	37.55	34.61	35.3	36.5	36.39	34.08	36.52	35.38	36.2	36.08	34.99	35.73	36.4	37.18	39.61	37.66	37.55	38.5	37.23	36.08	35.98	36.94
37	36.46	36.36	37.3	37.39	38.38	35.66	36.33	37.47	37.44	35.25	37.47	36.21	37.1	37.15	35.96	36.73	37.39	38.08	39.81	38.54	38.42	39.24	38.2	37.88	36.81	37.91
38	37.46	37.44	38.18	38.48	39.22	36.81	37.39	38.43	38.5	36.5	38.42	37.05	38.01	38.21	37	37.75	38.38	38.98	40	39.43	39.16	39.74	38.99	37.88	37.68	38.83
39	38.47	38.52	39.07	39.58	40.05	38.06	38.49	39.39	39.56	37.84	39.37	37.92	38.93	39.27	38.09	38.77	39.35	39.87	40.2	39.96	39.83	40.04	39.78	38.77	38.57	39.68
40	39.68	39.61	40.03	40.23	40.35	39.41	39.61	40.17	40.22	39.28	40.24	39.41	39.98	40.14	39.24	39.87	40.16	40.31	40.4	40.32	40.28	40.35	40.26	39.93	39.61	40.21



Gambar 7. Perbandingan Mean Awal dengan Mean Konversi IPA

(*equating*) skor peserta didik pada UASBN SD untuk mata ujian IPA di DI Yogyakarta dibandingkan skor peserta didik sebelum dilakukan penyetaraan (provinsi Jawa Barat sebagai provinsi acuan penyetaraan).

Garis lurus naik (yang berada di tengah) pada gambar adalah skor awal Jawa Barat. Garis lurus naik menunjukkan sebagai provinsi acuan Jawa Barat memiliki skor awal dan skor konversi sama. Sementara grafik skor konversi DI Yogyakarta



Gambar 8. Perbandingan Mean Awal dan Mean Konversi Propinsi DIY ke Jawa Barat

Tabel 4. Konversi Mata Pelajaran Ilmu Pengetahuan Alam (IPA)

JABAR	BALI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	SULBAF
1	1,07	1,47	1,38	1,21	1,38	1,58	1,48	1,46	1,07	1,53	1,59	1,33	1,35	1,23	1,35	1,31	1,13	1,56	1,38	1,59	1,45	1,36	1,3	1,3	1,42	1,41	1,55
2	1,96	2,62	2,47	2,19	2,46	2,8	2,81	2,6	1,96	2,72	2,81	2,38	2,42	2,21	2,42	2,35	2,05	2,76	2,47	2,81	2,58	2,43	2,34	2,34	2,53	2,51	2,75
3	2,84	3,77	3,56	3,16	3,55	4,02	4,02	3,73	2,84	3,91	4,03	3,56	3,44	3,21	3,49	3,39	2,97	3,97	3,56	4,04	3,72	3,5	3,37	3,65	3,61	3,95	3,95
4	3,72	4,92	4,64	4,14	4,63	5,24	5,24	4,87	3,72	5,1	5,26	4,49	4,55	4,2	4,56	4,44	3,89	5,17	4,64	5,26	4,85	4,57	4,41	4,76	4,72	5,16	5,16
5	4,61	6,06	5,73	5,11	5,72	6,46	6,46	6,01	4,6	6,29	6,48	5,54	5,62	5,19	5,62	5,48	4,8	6,38	5,73	6,49	5,98	5,65	5,44	5,88	5,82	6,36	6,36
6	5,49	7,21	6,82	6,09	6,8	7,69	7,69	7,15	5,49	7,48	7,7	6,6	6,69	6,18	6,69	6,52	5,72	7,58	6,82	7,71	7,12	6,72	6,48	6,99	6,93	7,56	7,56
7	6,37	8,36	7,9	7,06	7,89	8,91	8,91	8,29	6,37	8,67	8,93	7,65	7,76	7,17	7,76	7,56	6,64	8,79	7,91	8,94	8,25	7,79	7,51	8,1	8,03	8,76	8,76
8	7,26	9,51	8,99	8,04	8,97	10,13	10,13	9,43	7,25	9,86	10,15	8,7	8,82	8,16	8,83	8,6	7,56	9,99	8,99	10,16	9,39	8,86	8,54	9,22	9,14	9,96	9,96
9	8,14	10,66	10,08	9,01	10,06	11,35	11,35	10,57	8,13	11,05	11,37	9,76	9,89	9,15	9,89	9,64	8,48	11,2	10,08	11,39	10,52	9,93	9,58	10,31	10,24	11,16	11,16
10	9,02	11,8	11,17	9,99	11,15	12,57	12,57	11,7	9,02	12,24	12,6	10,81	10,96	10,44	10,96	10,68	9,4	12,4	11,17	12,62	11,65	11,01	10,61	11,44	11,34	12,36	12,36
11	9,91	12,95	12,25	10,96	12,23	13,79	13,79	12,84	9,9	13,43	13,82	11,86	12,03	11,13	12,03	11,72	10,32	13,61	12,26	13,84	12,79	12,08	11,65	12,46	12,45	13,4	13,4
12	10,79	14,1	13,34	11,94	13,32	15,01	15,01	13,98	10,78	14,62	15,04	12,92	13,09	12,11	13,1	12,77	11,24	14,81	13,35	15,07	13,83	13,13	12,65	13,48	13,55	14,44	14,44
13	11,67	15,24	14,43	12,91	14,4	15,92	15,92	15	11,66	15,71	15,94	13,98	14,12	13,1	14,17	13,81	12,16	15,84	14,45	16,02	14,87	14,18	13,65	14,5	14,66	15,42	15,42
14	12,56	16,27	15,46	13,89	15,5	16,82	16,82	15,95	12,55	16,59	16,85	15,04	15,14	14,09	15,15	14,85	13,08	16,86	15,57	16,97	15,85	15,23	14,65	15,5	15,78	16,34	16,34
15	13,44	17,31	16,48	14,86	16,64	17,73	17,73	16,89	13,43	17,48	17,76	16,14	16,14	15,08	16,09	15,87	13,99	17,87	16,71	17,91	16,83	16,29	15,64	16,47	16,92	17,26	17,26
16	14,32	18,34	17,5	15,89	17,77	18,63	18,63	17,83	14,31	18,37	18,67	17,23	17,14	16,29	17,02	16,88	15,04	18,88	17,86	18,85	17,82	17,36	16,64	17,44	18,06	18,18	18,18
17	15,24	19,38	18,52	16,92	18,91	19,54	19,54	18,77	15,21	19,25	19,58	18,33	18,14	17,94	17,96	17,9	16,22	19,89	19	19,79	18,8	18,42	17,64	18,41	19,2	19,09	19,09
18	16,44	20,41	19,54	17,96	20,05	20,44	20,44	19,71	16,36	20,14	20,49	19,42	19,14	19,51	18,89	18,92	17,41	20,89	20,14	20,73	19,78	19,49	18,63	19,38	20,33	20	20
19	17,54	21,44	20,56	18,99	21,19	21,34	21,03	20,65	17,51	21,03	21,4	20,51	20,14	20,98	19,83	19,94	18,59	21,9	21,27	21,66	20,76	20,56	19,63	20,35	21,45	20,91	20,91
20	18,64	22,46	21,58	20,02	22,32	22,23	21,91	21,59	18,67	21,91	22,31	21,59	21,13	22,35	20,77	20,96	19,78	22,89	22,4	22,6	21,74	21,62	20,63	21,31	22,55	21,81	21,81
21	19,74	23,48	22,6	21,05	23,45	23,12	23,22	22,52	19,82	22,9	23,22	22,67	22,13	23,61	21,7	21,97	20,96	23,88	23,53	23,53	22,72	22,69	21,62	22,28	23,63	22,71	22,71
22	20,83	24,5	23,61	22,07	24,57	24,01	24,13	23,44	20,97	23,69	24,13	23,73	23,12	24,78	22,64	22,99	22,14	24,86	24,65	24,46	23,7	23,75	22,62	23,24	24,68	23,6	23,6
23	21,93	25,5	24,61	23,09	25,68	24,88	24,57	24,36	22,12	24,57	25,03	24,78	24,11	25,88	23,57	24,01	23,32	25,83	25,77	25,38	24,67	24,81	23,62	24,15	25,7	24,48	24,48
24	23,03	26,5	25,61	24,09	26,77	25,76	25,76	25,27	23,26	25,46	25,94	25,82	25,1	26,92	24,5	25,03	24,49	26,79	26,88	26,3	25,65	25,87	24,62	25,14	26,69	25,36	25,36
25	24,13	27,49	26,6	25,08	27,85	26,62	26,85	26,17	24,4	26,34	26,85	26,85	26,08	27,98	25,44	26,05	25,66	27,73	27,98	27,22	26,63	26,92	25,61	26,08	27,64	26,24	26,24
26	25,22	28,47	27,58	26,06	28,9	27,48	27,47	27,07	25,53	27,23	27,76	27,86	27,06	28,3	26,37	27,06	26,81	28,66	29,07	28,13	27,96	27,96	26,61	27,02	28,57	27,11	27,11
27	26,32	29,44	28,55	27,02	29,92	28,34	28,33	27,95	26,66	28,12	28,67	28,86	28,02	29,83	27,3	28,08	27,95	29,58	30,16	29,03	28,59	28,99	27,61	27,96	29,46	27,99	27,99
28	27,41	30,41	29,51	27,97	30,92	29,19	29,12	28,83	27,77	29	29,58	29,84	28,99	30,73	28,23	29,1	29,08	30,48	31,23	29,93	28,57	30	28,6	28,88	30,33	28,87	28,87
29	28,5	31,36	30,45	28,89	31,89	30,03	30	29,7	28,88	29,89	30,49	30,81	29,94	31,5	29,16	30,12	30,2	31,37	32,29	30,83	30,55	31	29,6	29,8	31,17	29,76	29,76
30	29,58	32,3	31,39	29,81	32,84	30,88	30,38	30,56	29,97	30,77	31,4	31,76	30,89	32,45	30,09	31,14	31,3	32,25	33,34	31,73	31,53	31,98	30,6	30,71	32	30,66	30,66
31	30,66	33,23	32,31	30,7	33,76	31,73	31,53	31,42	31,06	31,65	32,31	32,69	31,83	33,29	31,01	32,15	32,39	33,11	34,38	32,62	32,51	32,94	31,59	31,61	32,81	31,58	31,58
32	31,73	34,15	33,23	31,58	34,65	32,58	32,38	32,27	32,13	32,54	33,22	33,62	32,76	34,11	31,94	33,17	33,47	33,98	35,41	33,52	33,48	33,88	32,59	32,51	33,6	32,51	32,51
33	32,8	35,07	34,14	32,45	35,52	33,42	33,32	33,12	33,19	33,42	34,13	34,54	33,69	34,91	32,86	34,19	34,53	34,84	36,44	34,42	34,46	34,81	33,59	33,4	34,39	33,47	33,47
34	33,87	35,98	35,04	33,31	36,36	34,28	34,38	33,98	34,28	34,13	35,03	35,44	34,61	35,7	33,79	35,21	35,59	35,69	37,46	35,32	35,44	35,72	34,59	34,21	35,17	34,44	34,44
35	34,93	36,88	35,94	34,17	37,19	35,13	35,13	34,83	35,28	35,18	35,94	36,35	35,53	36,49	34,71	36,23	36,64	36,54	38,07	36,22	36,42	36,63	35,58	35,18	35,94	35,42	35,42
36	35,98	37,78	36,84	35,02	37,96	35,98	35,88	35,69	36,32	36,06	36,85	37,25	36,44	37,28	35,63	37,24	37,69	37,4	38,51	37,12	37,28	37,53	36,58	36,06	36,71	36,41	36,41
37	37,04	38,51	37,5	35,87	38,58	36,84	36,74	36,54	37,36	36,94	37,76	38,11	37,35	38,04	36,55	38,07	38,35	38,21	38,95	37,95	38,07	38,24	37,58	36,95	37,49	37,4	37,4
38	38,08	39,08	38,54	36,73	39,13	37,7	37,69	37,4	38,36	37,82	38,56	38,79	38,26	38,74	37,48	38,76	38,97	38,86	39,39	38,68	38,77	38,89	38,44	37,81	38,34	38,31	38,31
39	39,05	39,65	39,33	37,58	39,68	38,78	38,78	38,55	39,72	38,89	39,33	39,47	39,16	39,44	38,64	39,46	39,58	39,52	39,84	39,41	39,46	39,53	39,26	38,88	39,2	39,19	39,19
40	40,02	40,22	40,11	38,86	40,23	39,93	39,9	39,85	40,07	39,96	40,11	40,16	40,05	40,15	39,88	40,15	40,19	40,17	40,28	40,14	40,15	40,18	40,09	39,96	40,07	40,06	40,06

(dalam grafik ditunjukkan sebagai garis di bagian atas) pada bagian skor awal 0 sampai dengan skor 40 berada di atas garis skor propinsi acuan (Jawa Barat). Hal ini menunjukkan bahwa setelah dilakukan penyetaraan skor peserta didik DI Yogyakarta terhadap skor peserta didik Jawa Barat, menunjukkan bahwa kemampuan peserta didik di DI Yogyakarta berada di atas kemampuan peserta didik di Jawa Barat. Hasil analisis dalam bentuk tabel konversi mata pelajaran Ilmu Pengetahuan Alam (IPA) untuk skor-skor di seluruh propinsi yang di *equate* ke Jawa Barat sebagai propinsi acuan dapat dilihat pada tabel 4.

Pada tabel 4 terjadi perubahan skor setelah dikonversi, baik menunjukkan kenaikan maupun penurunan. Terdapat 7 propinsi yang mengalami penurunan skor setelah dikonversi yaitu Bengkulu, DKI Jakarta 1, DKI Jakarta 2, Gorontalo, Jawa Tengah, Kalimantan Timur, Papua, dan Sulawesi Tenggara.

Penurunan skor tersebut menunjukkan bahwa soal-soal di propinsi tersebut tingkat kesukarannya lebih rendah dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat). Sementara pada 26 propinsi lainnya menunjukkan kenaikan skor, yang menggambarkan bahwa tingkat kesukaran soal-soal di propinsi tersebut lebih tinggi dibandingkan dengan tingkat kesukaran soal di propinsi acuan (Jawa Barat).

## Simpulan dan Saran

### Simpulan

Penyetaraan (*equating*) UASBN SD tahun 2009/2010 yang dilakukan untuk setiap mata pelajaran menggunakan propinsi Jawa Barat sebagai acuan, pemilihan Jawa Barat sebagai propinsi acuan didasarkan pada data hasil analisis item dengan menggunakan *Item Response Theory* yang menunjukkan data dari Jawa Barat masuk kategori stabil.

Berdasarkan hasil analisis penyetaraan (*equating*) UASBN SD tahun 2009/2010 pada

setiap mata pelajaran dapat disimpulkan sebagai berikut. Pertama, analisis penyetaraan menunjukkan bahwa hasil konversi antarpaket yang digunakan di seluruh wilayah Indonesia, dengan menggunakan soal *anchor* yang sama untuk setiap pakatnya menghasilkan nilai konversi yang bervariasi. Ada yang menghasilkan skor konversi tetap, kenaikan skor, maupun penurunan skor. Kedua, hasil *equating* (hampir pada seluruh propinsi) menunjukkan bahwa tingkat kesukaran soal selain *anchor* pada setiap propinsi berbeda. Soal-soal daerah cenderung lebih mudah dibandingkan soal nasional (soal *anchor*). Ketiga, sebagian besar nilai pada setiap mata pelajaran di daerah (propinsi) mengalami perubahan yang cukup signifikan setelah dilakukan konversi. Keempat, banyak faktor yang menyebabkan penurunan maupun kenaikan skor, diantaranya penggunaan soal yang tidak standar, pengadministrasian tes yang tidak standar, kemampuan peserta ujian, kondisi psikologis siswa, kemampuan guru dalam menulis soal, kompetensi mengajar guru, dan sebagainya.

### Saran

Mengacu pada simpulan, maka disarankan agar: 1) Perlu ditingkatkan kualitas soal-soal yang disusun di provinsi sehingga setara dengan soal-soal nasional dengan melakukan pelatihan penulisan soal bagi guru-guru penulis soal; 2) Perlu dipertimbangkan untuk membentuk Bank Soal Daerah sehingga soal-soal yang kelak digunakan untuk UASBN merupakan soal-soal yang standar (memiliki karakteristik soal); dan 3) Pengadministrasian tes harus standar untuk menghindari adanya kecurangan-kecurangan saat tes berlangsung. Pengawasan yang seharusnya dilakukan oleh pengawas ruang ujian, dipastikan sudah dilakukan sesuai POS (Prosedur Operasional Standar).

### **Pustaka Acuan**

- Angoff, W.H., 1971. Scale, norms, and equivalent scores, In RL Thorndike (ed), *Educational Measurement*, Washington DC: American Council on Education.
- Crocker, L., Algina, J., 1986. *Introduction to Classical and Modern Test Theory*, NY: Holt, Rinehart and Winston, Inc.
- Grondlund, E.N., 1982. *Constructing Achievement Test*, EC: Prentice Hall, Inc.
- Hamblethton, R.K., Swaminathan H.J., 1985. *Item Response Theory : Priciples and Applications*, Boston, MA: Kluwer Academic Publisher.
- Hamblethton, R.K., Swaminathan H.J., Roger, H.J., 1991. *Foundations of Item Response Theory*, London: Sage Publications.
- Kolen. 2004. *Software Common Item Program for Equating (CIPE)* versi 2.0.
- Kolen, 1984, Effectiveness of Analysis in Equipercetile Equating, *Journal of Educational Statistic*, 9, pp.25-44
- Kolen, M. J., & Brennan, R. L., 2004. *Test equating, scaling, and linking: Methods and practices*. New York: Springer.
- Livingstone, 2004, *Equating Test Scores (Without IRT)*, Princeston, Nj: ETS
- Petersen, N.S. 1989. Educational Measurement, Scaling, Norming, and Equating, in R.L Linn (ed), *Educational Measurement*, NY: Macmillan.
- Suryabrata, Sumadi., 1987. *Pengembangan Tes Hasil Belajar*, Jakarta: CV. Rajawali
- Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional beserta penjelasannya.
- Weiss, D.J., 1983. *New Harizons in testing : A Test of adequacy of curvilinear score equating models*, NY: Academic Press.